Routledge
Taylor & Francis Group

Research Papers in Education

ISSN: 0267-1522 (Print) 1470-1146 (Online) Journal homepage: http://www.tandfonline.com/loi/rred20

Evaluating teachers' professional development initiatives: towards an extended evaluative framework

Emmelien Merchie, Melissa Tuytens, Geert Devos & Ruben Vanderlinde

To cite this article: Emmelien Merchie, Melissa Tuytens, Geert Devos & Ruben Vanderlinde (2016): Evaluating teachers' professional development initiatives: towards an extended evaluative framework, Research Papers in Education, DOI: 10.1080/02671522.2016.1271003

To link to this article: http://dx.doi.org/10.1080/02671522.2016.1271003

RESEARCH PAPERS IN EDUCATION, 2016
http://dx.doi.org/10.1080/02671522.2016.1271003

Routledge
Taylor & Francis Group

Evaluating teachers' professional development initiatives: towards an extended evaluative framework

Emmelien Merchie, Melissa Tuytens, Geert Devos and Ruben Vanderlinde

Department of Educational Studies, Ghent University, Ghent, Belgium

ABSTRACT

Evaluating teachers' professional development initiatives (PDI) is one of the main challenges for the teacher professionalisation field. Although different studies have focused on the effectiveness of PDI, the obtained effects and evaluative methods have been found to be widely divergent. By means of a narrative review, this study provides an extended framework to guide the evaluation of (the effectiveness of) PDI. Furthermore, and in accordance with this framework, an overview of measurement instruments that are currently in use is provided. The study concludes with implications for educational research and developers of PDI.

ARTICLE HISTORY
Received 4 April 2016
Accepted 27 November 2016

KEYWORDS
Teacher professional development initiatives; evaluation; extended framework; measurement instruments

Introduction

Teachers play a crucial role in teaching and learning. However, they face the challenge to keep abreast with the rapidly growing knowledge base in education (Borko 2004; Desimone 2009; Guskey 2000). As such, the need for their professional development (PD) throughout their school career is frequently stressed in educational improvement plans, empirical research studies and meta-analyses (e.g. Guskey 2003; Hattie 2009). To address teachers' PD needs, various professional development initiatives (PDI) have already been developed. However, the evaluation of these PDI and their achieved effects has been found to be widely divergent. Therefore, and not surprisingly, the National Research Council of the US (2010) considers mapping the effectiveness of professional developments as one of the main challenges for the educational research field. Still, until now, little clarity exists on how to evaluate professional development (King 2014). More particularly, questions arise as to which outcomes are appropriate measures of 'qualitative' teaching and the possible benefits of this teaching for student learning (Guskey 2000, 2014). Furthermore, how these outcomes can be evaluated and measured in a focused and systematic way is also questioned (Guskey 2000; Muijs and Lindsay 2008). Although previous research has mapped the possible effects of PD (e.g. Desimone 2009; Van Veen, Zwart, and Meirink 2012) and evaluation models of PD have been developed (e.g. Guskey 2000; Muijs and Lindsay 2008), research on these topics is fragmented and an integrated view on PD evaluation is still missing. Consequently, researchers appeal for an extended framework to make thoughtful and informed decisions about the evaluation of high-quality PDI (Borko 2004; Desimone 2009; Guskey 2000; King 2014). The current study aims to answer this call by interweaving research on PD outcomes on the one hand, and PD evaluation models on the other, resulting in an extended evaluative framework that can be used in research and practice.

CONTACT Ruben Vanderlinde Ruben.Vanderlinde@UGent.be

Evaluating PDI

Definition of PDI

Professional development can entail a variety and multiplicity of initiatives (e.g. workshops, seminars, conferences, summer institutes, action research and so on) (Desimone 2009), leading to various definitions of PDI in the literature. These definitions reflect either narrow or more broad views on professional development and PDI (e.g. Avalos 2011; Borko 2004; OECD 2010; Schachter 2015; Timperley 2008), which makes it essential to carefully delineate what PDI exactly entail. In this study, PDI are defined as 'activities explicitly designed for and provided to educators or certified educational professionals with a focus on enhancing their own and their students' knowledge, skills, and attitudes' (based on Guskey 2000; Tienken and Achilles 2003). This definition reflects important components that PDI can affect, such as teachers' knowledge, skills and attitudes as well as students' learning. Furthermore, the definition explicitly stresses the thoughtful design of PDI with a systematic, intentional and ongoing effort (Guskey 2000). By focusing on the explicit design of PDI, more informal learning activities in which teachers participate are not considered in this study. However, their value and their immixture with formal activities are truly acknowledged and described in the research literature (e.g. De Neve, Devos, and Tuytens 2015; Desimone 2009; Grosemans et al. 2015; Richter et al. 2011; Vanblaere and Devos 2016).

Conceptual framework

Evaluating PDI is important for various (research) goals, such as gaining a better understanding of a PDI's dynamic nature, initiating positive change and improvement, better informing and guiding reform efforts and the increasing pressure for accountability (Guskey 2014). In this respect, the evaluation of PDI serves different – mostly blended – purposes. More specifically, a PDI's evaluation can be applied for planning (i.e. the appraisal of a PDI's critical attributes), for formative evaluation (i.e. providing information during the PDI as to whether things go as planned and progress is made) or for summative evaluation (i.e. providing judgements on the programme's overall merit after PDI completion) (Guskey 2000, 2014). A first important step in evaluating a PDI is determining which outcomes are aspired to and which outcomes can be expected after following a PDI. In this respect, several conceptual models have been proposed to map these outcomes (e.g. Desimone 2009; Guskey 2000; Wallace 2009). For example, Guskey and Sparks (as cited in Guskey 2000) proposed a model with two key components: the quality of the professional development and improved student outcomes. The relationships between these two components are mediated by administrator, teacher, parent and school characteristics. A more recent, frequently cited model is proposed by Desimone (2009). According to her, professional development comprises the following steps: (1) teachers experience effective professional development, (2) the professional development increases teachers' knowledge and skills and/or changes their attitudes and beliefs, (3) teachers use their new knowledge and skills, attitudes and beliefs to improve the content of their instruction or their approach to pedagogy, or both and (4) the instructional changes foster increased student learning. All these steps are embedded within a context comprising teacher and student characteristics, curriculum, school leadership and the policy environment (Desimone 2009). Parallels can be drawn between Guskey's (2000) and Desimone's (2009) models as they both acknowledge PD characteristics, qualitative teaching and student learning and context as important elements. However, the framework of Desimone (2009) has an important asset as it explicitly portrays interactive relationships between (a) features of professional development, (b) increased teacher knowledge and skills, and changes in their attitudes and beliefs, (c) change in instruction and (d) improved student learning (Figure 1). This means, for example, that a change in teachers' beliefs can lead to a change in practice or inversely. Other researchers have also applied or adapted Desimone's model (2009) in their research (e.g. Boston 2013; Kang, Cha, and Ha 2013; Van Veen, Zwart, and Meirink 2012).

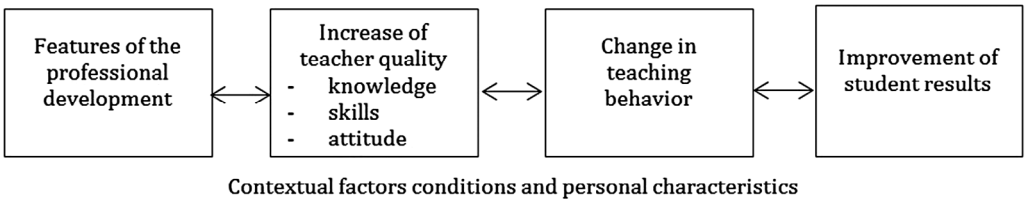

Figure 1. Teachers' professional development framework (based on Desimone 2009; Van Veen, Zwart, and Meirink 2012).

Table 1. Guskey's five stages in evaluating professional learning initiatives (adjusted from Guskey 2000).

Evaluation stage	Selection of questions addressed
1. Participants' reactions	Did they like it? Was their time well spent? Did the material make sense? Was the leader knowledgeable and helpful? Was the room the right temperature? Were the chairs comfortable?
2. Participants' learning	Did participants acquire the intended knowledge and skills?
3. Organisational support and change	Were sufficient resources made available? Were problems addressed quickly and efficiently? Was implementation advocated, facilitated and supported? Was the support public and overt? What was the impact on the organisation?
4. Participants' use of new knowledge and skills	Did participants effectively apply the new knowledge and skills?
5. Student learning outcomes	What was the impact on students? Did it influence students' physical or emotional well-being? Are students more confident as learners? Is student attendance improving?

Figure 1 illustrates that the outcomes of PDI are not straightforward since there is a complex and dynamic interplay between different components. Within this dynamic interplay, it is important to consider three different relationships when evaluating PDI (Desimone 2009; Van Veen, Zwart, and Meirink 2012; Wayne et al. 2008). First, it is important to consider the links between the professional development initiative's features and changes in teacher knowledge, skills, attitudes and/or instruction. A second important set of relationships refers to those between the features of PDI, changes in teacher knowledge, skills, attitudes, instruction and changes in student outcomes. Third, contextual factors and personal characteristics are also important to consider when implementing and sustaining successful PDI (Desimone 2009; Van Veen, Zwart, and Meirink 2012). This general model (Figure 1) might offer guidance to evaluators of PDI and will serve as a starting point in this study. More particularly, each component in this general framework will be extended.

Evaluation models

As described in the previous section, researchers have already focused on possible outcomes of a PDI and the relationship between these outcomes. Others have focused on how these outcomes should be evaluated and they describe in more general terms sequences or ways to undertake this PDI evaluation. In this respect, specific evaluation models have been developed in the literature (e.g. Guskey 2000, 2014; Hammond 1973; King 2014; Muijs and Lindsay 2008; Stake 1967). Those models generally have in common that they draw attention to evaluating particular factors before, during and after the PDI, as well as taking into account contextual factors that might possibly influence a PDI's effectiveness. The most frequently cited evaluation model is Guskey's (2000, 2014) adoption of Kirkpatrick's model (1994) for evaluating training programmes. He describes five hierarchically arranged professional learning 'evaluation stages': (1) participants' reactions, (2) participants' learning, (3) organisational support and change, (4) participants' use of new knowledge and skills and (5) student learning outcomes (Table 1). Evaluation at all levels is important and for every level Guskey (2000) describes questions to be addressed, how information can be gathered, which outcomes are assessed and measured and

how information can be used. When progressing through these five stages, Guskey (2000) states that data gathering requires more time and resources. Despite the importance of focusing on these five different stages in the evaluation of professional development, researchers caution for the inadequate and ineffective evaluation of professional development which might be too shallow (e.g. are participants enjoying the experience?) or too brief (e.g. gathering quick and ample evidence) (Guskey 2000; Muijs and Lindsay 2008). Although a positive trend is shown from these low-level evaluations (e.g. small-scale investigations) to more high-level evaluations (e.g. comparison of programme features; using objective measures) (Hill, Beisiegel, and Jacob 2013), the extent to which all stages are considered in the evaluation of PDI remains unclear.

In sum, the evaluation of PDI serves different purposes associated with planning PDI and its formative and summative evaluation. Previous studies illustrate the importance of undertaking this evaluation in a focused and systematic way (Guskey 2000; Muijs and Lindsay 2008). In a focused way, the conceptual model proposed above (see Figure 1) portrays the complex and dynamic interplay between different PDI outcomes and, in this respect, sheds a light on *what* can be evaluated in PDI. In a systematic way, evaluation can be undertaken by following different evaluation stages showing us *how* we can evaluate PDI. However, two main challenges arise when considering these aspect in the current research literature, namely: (1) the lack of a more detailed and fine-grained framework of possible PDI outcomes to conduct PDI evaluation in a focused way and (2) information on how to evaluate these different PDI outcomes in a systematic way using various measurement instruments. These are described in the next section.

Conceptual and methodological challenges in evaluating PDI

A first conceptual challenge lies in what to evaluate when considering the effectiveness of a professional development initiative. A second methodological challenge concerns how to undertake this evaluation in a systematic way.

First, mapping the effects of a PDI is an important though complex and challenging endeavour. This endeavour is described by the National Research Council of the US (2010) as a major challenge for educational researchers. The model described in Figure 1 represents a starting point for addressing this challenge as it portrays various components and relationships to consider in evaluating a PDI's effectiveness. However, the components included in the model are described rather vaguely. Because of this, it remains unclear which particular effects can actually be expected within each of these components when teachers participate in PDI. Furthermore, little information is available on whether impact studies already consider different components in the model when evaluating PDI. Desimone (2009) acknowledges this restriction by stating that important elements might not yet be included in her model due to the lack of impact research. Therefore, a more detailed and fine-grained investigation of possible PDI outcomes is warranted, implying an extension to the model of Desimone as presented in Figure 1. In this way, a more differentiated view on possible outcomes that are important in the evaluation of PDI is provided.

A second methodological challenge relates to how these outcomes are evaluated and/or measured. This is important, as the evaluation of high-quality PDI requires different types of inquiries and a consistent set of research instruments (Borko 2004; Desimone 2009). Although Desimone (2009) suggests applying her model for evaluation, few direct guidelines are provided. Guskey (2000, 2014) does provide these guidelines by proposing a sequential evaluation at multiple levels. As can be noticed, parallels can be drawn between his evaluation stages (Table 1) and Desimone's (2009) framework (Figure 1). More particularly, Guskey's first two levels (participants' reactions and learning) resemble the second component in Desimone's framework 'teacher quality', the third level (organisational support and change) aligns with 'school organisational conditions', level 4 (participants' use of new knowledge and skills) fits within the third component 'changing teacher behaviour' and level 5 (student learning outcomes) focuses on 'improvement of student results'. However, some dissimilarity is shown between Guskey's evaluation stages and more recent insights on the effectiveness of PDI. For

instance, an important element lacking in Guskey's five-stage evaluation model is the evaluation of a PDI's effective features. These are crucial considerations, as different results can be expected from different PDI, taking into account whether the PDI meets all, some or none of these effective features. Furthermore, Guskey represents the evaluation of PDI as linear, hierarchical stages, which contrasts with the complex interplay of possible PDI outcomes (Earley and Porritt 2014). This study aims to merge current insights on important outcomes in PDI and research on the systematic evaluation of PDI.

In sum, this study specifically focuses on a conceptual and methodological research aim:

(1) Conceptually, this study aims to establish an extended framework for the evaluation of PDI, representing major categories of PDI outcomes that should be taken into account when evaluating a PDI's effectiveness.
(2) Methodologically, this study aims to provide a consistent set of methods and instruments to evaluate the components integrated into the evaluative PDI framework (i.e. how can these effects be evaluated?) and guidelines on how to undertake this evaluative endeavour in a focused and systematic way.

Method

To answer the research aims, a systematic narrative synthesis was undertaken, referring to 'an approach to the systematic review and synthesis of findings from multiple studies that relies primarily on the use of words and text to summarise and explain the findings of the synthesis' (Popay et al. 2006, 5). This method, also applied in previous research (e.g. Baeten and Simons 2014; Dochy, Segers, and Buehl 1999), leads to conclusions which are drawn into a comprehensive en coherent interpretation contributed by the researchers' experience and existing theories (Kirkevold 1997; Popay et al. 2006).

Search procedure and inclusion criteria

A systematic literature search was conducted for the years 2000–2015 in international databases (i.e. ISI Web of Knowledge, SSCI, EBSCO/ERIC, ICO-journals and Google Scholar) and specialist libraries (i.e. OECD and Eurydice) following a three-step method (Dochy, Segers, and Buehl 1999). In a first search, the following search terms were used: 'teacher professional development', 'teacher learning', 'evaluating professional development', 'evaluation model', 'assessing professional development', 'effects of professional development' and 'effectiveness of professional development'. In a second search, these general terms were combined with more detailed keywords referring to specific PDI (e.g. 'workshop', 'conference', 'action research', 'teacher induction programmes' and so on) or studied variables (e.g. 'teacher motivation' and/or 'quality of trainer'). Third, relevant references in previously selected articles were screened. The following quality criteria were used for inclusion in the analysis:

(1) Empirical peer-reviewed work, published in a journal, dissertation or report commissioned by a governmental institute or agency.
(2) A clear description of the professional development initiative under investigation to verify its concordance with our previously described definition of PDI as 'activities explicitly designed for and provided to educators or certified educational professionals with a focus on enhancing their own and their students' knowledge, skills, and attitudes' (based on Guskey 2000; Tienken and Achilles 2003).
(3) The PDI has to be conducted in compulsory education (i.e. primary or secondary education).
(4) A clear description of the outcome measures, enabling these measures to be positioned in Desimone's (2009) conceptual framework.
(5) A well-elaborated and transparent method, entailing a clear description of the evaluative procedure used (methodology, measurement method, instruments, information on the reliability of the measures, data collection, overview of different steps in the data-analysis, etc.).

The initial search yielded 79 manuscripts. After matching these studies to the inclusion criteria mentioned above, 68% or 54 manuscripts in total were included in the analysis (see Appendix).

Coding procedure and synthesis

Information on all selected studies was collected in tables and coded according to the following aspects: authors, year published, country, publication type (i.e. general effect study considering several PDI, specific effect study considering a particular PDI and/or meta-analysis), level of education (i.e. primary or secondary), participants (type and/or number), subject domain, design, specific PDI, effect measures, effect measures situated in Desimone's model (2009), measurement instruments, inclusion of measurement instruments in the appendix, data analysis, main research questions and conclusions. To condense these extensive tables, only information directly relevant to postulated research questions is included in the appendix, that is information on the studied PDI outcomes, the applied measurement instruments and whether or not the measurement instruments are included in the appendix of the study.

Results

A first aim of this study was to extend Desimone's model (2009) by positioning important PDI aspects and outcomes within an extended evaluative framework. Next, in response to our second research aim, an overview is provided of a consistent set of measurement methods and instruments to undertake the evaluation of PDI in a focused systematic way.

Extended evaluative framework

In what follows, each of the components in Desimone's framework will be successively elaborated in greater detail. More specifically, for each of the components, subcategories are specified that are empirically found to be important in the evaluation of PDI. In what follows, our extended evaluative framework is presented.

Key features of professional development

Various studies meeting our quality criteria (see 10–12) have focused on identifying key features of successful PDI, the first component in Desimone's (2009) framework. When overviewing research in this area, eight main features are repeatedly proven to be effective and are consistently noted in research (Desimone 2009; Hammerness et al. 2005; Kang, Cha, and Ha 2013; Kedzior 2004; Postholm 2012; Timperley et al. 2007; Van Veen, Zwart, and Meirink 2012; Wayne et al. 2008). Additionally, a ninth important feature was identified in more recent impact studies (Borko 2004; Cheng and So 2012; Morrison 2014; Van den Bergh, Ros, and Beijaard 2015; Walker et al. 2012). These nine features can be subdivided into core and structural features (Desimone 2009; Garet et al. 2001). Whereas core features refer to the substance of the PDI, structural features refer to characteristics of the activities' structure or design. In 31% of the studies included in our review, one or more of these effective intervention features are explicitly addressed or evaluated when considering the effectiveness of a particular professional development initiative.

Core features

(1) *Content focus*: focus on student learning and informed by evidence on student learning;
(2) *Pedagogical knowledge*: focus on enhancing the knowledge and skills to teach in these content areas, which also means considering students' prior knowledge;
(3) *Coherent and evidence-based*: experiences in the PDI should be aligned with teachers' goals, standards and current reforms, and informed by theory and meaningful research evidence;

(4) *Ownership*: responding to teachers' self-identified needs and interests. PDI are more meaningful to teachers when they exercise ownership of its content and process;

Structural features

(5) *Duration*: extended and intensive programmes. Although no exact tipping point exists, research supports activities that are spread out (e.g. courses with a follow-up during a semester), including 20 h of contact time or more (Desimone 2009);
(6) *Collective or collaborative participation*: through collaboration with internal and external peers (e.g. observing each other's practices and giving feedback);
(7) *School or site based*: incorporated into teachers' daily work;
(8) *Active learning*: inquiry-based through continuous inquiry of practice and reflection on professional and academic knowledge. Less resistance to professionalisation is experienced by teachers when they are co-creators, rather than consumers of knowledge;
(9) *Trainer quality*: in addition to the eight features mentioned above, several recent studies highlight an additional ninth important feature of PDI, that is the quality of the trainer (Borko 2004; Cheng and So 2012; Morrison 2014; Van den Bergh, Ros, and Beijaard 2015; Walker et al. 2012). The trainer can be regarded as the 'facilitator', guiding teachers as they construct new knowledge and practices (Borko 2004). In this respect, both the trainers' (content) knowledge and skills (e.g. supporting self-regulation; providing qualitative feedback) seem to be important. For example, concerning the trainers' content knowledge, Morrison (2014) attributes the success of the studied intervention to the collaboration with scientists or engineers. These are content knowledge experts in their field, providing teachers with better understandings. Van den Bergh, Ros, and Beijaard (2015) underline the importance of trainers' skills that can play a crucial role in a PDI' effectiveness. In their case study, the trainer's feedback played an important role in compensating primary school teachers' lack of self-regulation. They state that trainers should be able to give tailored feedback to each individual teacher, addressing their concerns, practices and learning characteristics. Other researchers also point to the importance of receiving specific, constructive (one-on-one) feedback provided by the trainer (Cheng and So 2012; Morrison 2014). Walker et al. (2012) point to considering the skill of workshop leaders in professional development. Thus, the trainers' quality seems to be an additional important aspect in studying the initial features of a PDI.

Increased quality

In the previous section, effective intervention features were unravelled which are important to consider in PDI evaluation. A second component to be unravelled includes changes in teacher quality, which is considered as a PDI outcome in 57% of the studies included in our analysis based on the applied quality criteria (see 10–12). Desimone (2009) breaks down teacher quality into knowledge, skills and attitudes, which resembles Guskey's (2000) distinction between cognitive, psychomotor and affective goals of participants' learning. In what follows, we investigate whether these main categories are reflected in the outcome measures of the selected studies and whether these categories can be further extended.

Cognitive goals are related to teachers' content and pedagogical knowledge. In this respect, Shulman's (1986) distinction between teachers' (subject matter) content knowledge, pedagogical content knowledge and curricular knowledge is frequently referred to. First, subject matter content knowledge refers to the amount and organisation of knowledge in the teachers' mind. Several authors have found positive effects of a PDI on teachers' content knowledge, for instance, in science (e.g. Buczynski and Hansen 2010), language (e.g. Goldschmidt and Phelps 2010), mathematics (e.g. Garet et al. 2001) or citizenship education (Willemse et al. 2015). Second, pedagogical content knowledge (PCK) includes a teachers' armamentarium of instructional strategies, understanding of what makes learning easy or difficult and how students develop insights in specific subject matter (Shulman, 1986; Van Driel and Berry 2012).

Some authors additionally separate PCK into 'knowledge of teaching' and 'knowledge of students' (Levenson and Gal 2013; Loewenberg Ball, Thames, and Phelps 2008). For instance, teachers' awareness of special need students might fall into this last subcategory. In this respect, Levenson and Gal (2013) found a positive intervention effect on a teacher's awareness of the cognitive and affective needs of mathematically talented students. The study of Cheng and So (2012) might fit into the 'knowledge of teaching' subcategory as they found a positive relationship between the PDI and increasing teachers' PCK in inquiry learning. Some authors also distinguish 'technological pedagogical content knowledge' as a distinct category when PCK is technology related (Abuhmaid 2011; Mishra and Koehler 2006; Niess 2005). Third, curricular knowledge refers to knowledge of curricular programmes and instructional materials designed for teaching particular subjects (Shulman 1986). However, this knowledge was not found to be investigated in our selected impact studies.

Skills relate to what participants are able to do with what they have learnt during PDI (Guskey 2000). Within the component of 'teacher quality', these skills have to be explicitly distinguished from teachers' actual classroom behaviour. In most instances, teachers are asked to estimate their skill level, or their ability to implement specific strategies is tested in simulations (Guskey 2000). Very few studies in our review evaluate teachers' skills as part of teachers' quality. An exception was found in Vogt and Rogalla's (2009) study. They assessed teachers' adaptive planning competency (i.e. ability to adjust lesson planning to diverse student needs), and found positive effects of a PDI on teachers' competency. These authors used vignettes to assess teachers' skills without observing this specific competency in an actual classroom setting. Also micro-teaching could be regarded as a particular PDI whereby a simulation is used to assess teachers' skill level.

As to the *affective goals,* research investigating 'teachers' attitudes or beliefs about teaching and learning' and 'beliefs about themselves' as a PDI outcome measures can be discerned. Concerning beliefs about learning and teaching, for instance, De Vries, van de Grift, and Jansen (2014) distinguished three distinct belief profiles among Dutch secondary school teachers (i.e. student oriented, subject matter oriented and combined subject matter and student oriented). More specifically, the more student and subject matter oriented teachers are, the higher their participation in continuous professional development. Other studies report positive changes in beliefs on learning (Bakkenes, Vermunt, and Wubbels 2010; James and McCormick 2009), inquiry-based practices (McCutchen et al. 2002) or recognising the potential of using mobile phones in teaching (Ekanayake and Wishart 2014) after a PDI. When it comes to teachers' beliefs about themselves, Morrison (2014) found an increase in elementary school teachers' self-efficacy to implement science inquiry in their classroom. In this respect, Levenson and Gal (2013) underline that 'a change in classroom practice may not take place without the teacher believing that he or she is capable of affecting this change' (Levenson and Gal 2013, 1109). Fishman et al. (2003) underline the immixture of teachers' knowledge and beliefs.

Change in instruction

In 74% of the studies included in our analysis, changes in teacher instruction after a professional development initiative are investigated, which is the third component in Desimone's model. Generally, two broad categories of outcome measures emerged from our analysis to be classified as either 'changes in instructional strategies or practices' or in 'changes in interaction patterns' (between teachers and students, among teachers and among students). In what follows, each subcategory is illustrated with some specific study results. For instance, for *instructional strategies and practices,* researchers found that a PDI focusing on specific instructional practices (i.e. use of technology, higher order instructional methods and alternative assessment practices (Desimone et al. 2002); use of sociocultural instructional practices (Teemant, Wink, and Tyra 2011) led to an increased use of those practices in the classroom. Garet and colleagues found a positive effect of their PDI on teachers' use of research-based instructional practices (Garet et al. 2008). Bakkenes and colleagues found six categories of change in instruction, that is experimenting, considering own practice, experiencing friction, struggling not to revert to old ways, getting ideas from others and avoiding learning after following a national innovation programme (Bakkenes, Vermunt, and Wubbels 2010). Others have focused on investigating changes in *interaction*

patterns. For instance, Butler et al. (2004) report the positive influence of a PDI on teachers' ability to match instruction to student needs. Kiemer et al. (2015) found positive effects of a video-based teacher professional development initiative focusing on providing constructive feedback on teacher–student interactions. Voerman et al. (2015) found positive effects of their multicomponent PDI on teachers' learning-enhancing feedback behaviour. Also Consuegra (2015) studied the effects of a PDI on teacher–student interaction but did not find significant changes in the PDI focusing on teachers' feedback patterns to boys and girls. Chamberlin (2005), on the other hand, focused on interaction patterns among teachers, finding a positive relationship of the PDI on these interaction patterns when synthesising students' solution strategies. When considering the interaction patterns among students, the study of Mikami et al. (2011) reports on more positive classroom peer interactions as a results of a teacher PDI.

Improved student learning

Less than half of our selected articles (40%) address improvements of student results as an outcome measure of a professional development initiative, the fourth component in Desimone's conceptual framework (2009). In almost all studies, student learning outcomes are considered as outcome measures of PD alongside other areas of improvement related to teacher knowledge of teachers' change in instruction. As to student learning outcomes, a distinction can be made between domain-specific and domain-general outcomes. On the one hand, studies have found positive effects of teachers' professional development on increasing students' language and literacy achievement (Garet et al. 2008; McCutchen et al. 2002; Wasik and Hindman 2011), science achievement (Morge, Toczek, and Chakroun 2010) or mathematics achievement (Antoniou and Kyriakides 2013; Saxe, Gearhart, and Nasir 2001). These studies all include measures of students' domain-specific, subject-related knowledge and skills. On the other hand, other researchers focused on stimulating more domain-general skills, such as students' social–emotional development (Bierman et al. 2008), and self-regulated learning (e.g. students' higher order learning and use of cognitive elaboration strategies (Pehmer, Gröschner, and Seidel 2015)).

School context and personal characteristics

In 9% of the articles, the role of contextual factors in PDI is explicitly referred to. To unravel this component, we refer to Avalos' (2011) distinction between macro-societal conditions and the micro-context (school culture). Most studies in the analysis document the role of the *micro-context* or *school culture* (i.e. administrative and organisational structures; atmosphere). Ingvarson, Meiers, and Beavis (2005) point to the importance of the professional community as a mediating variable in enhancing teachers' knowledge and practice. Buczynski and Hansen (2010) found that a lack of resources (e.g. lack of science kits for teaching; technology) formed a barrier for teacher's participation in PDI. Additionally, important micro-contextual factors were the coordination of a PDI across school personnel (Butler et al. 2004), school management practices (James and McCormick 2009) and the principal's attitude towards the PDI (Liu 2013). Postholm (2012) highlights that the teachers' PDI may be influenced by a positive school culture with a good atmosphere. A school's socio-economic status (as measured by free or reduced school lunches) also seems to relate to teachers' PDI. In this respect, Supovitz and Turner (2000) found that the school's socio-economic status related to changes in practice more than the principal's supportiveness. Opfer and Pedder (2011) found that teachers in high achieving schools engaged in more active, collaborative activities of longer duration. Under *macro-conditions*, Avalos includes 'the nature and operation of educational systems, policy environments and reforms, teacher working conditions and history facts that determine what is accepted or not as suitable forms of professional development' (Avalos 2011, 12). In this respect, Abuhmaid (2011) describes PDI in Jordan, where a lack of national ICT standards led to a fragmentation of PDI. It is highlighted in the literature that other macro-conditions might also play a role in the PDI of teachers, such as whether following PDI is part of a teachers' professional duty (OECD 2010) or whether PDI are top-down directed from the educational government which might hinder teachers' enthusiasm (King 2014; Nir and Bogler 2008).

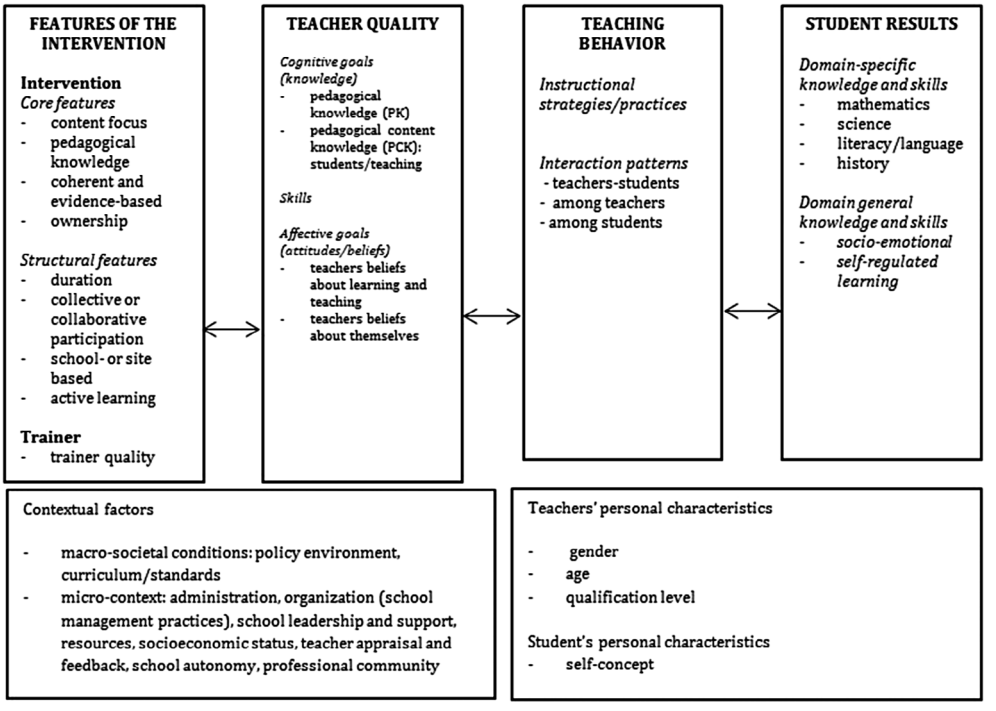

Figure 2. Extended evaluative framework for mapping the effects of professional development initiatives (PDI).

In addition to contextual factors, teacher and student characteristics can also affect a PDI's effectiveness. In respect of teacher characteristics, the OECD (2010) identified gender, age and teachers' qualification level as factors related to teachers' participation in PDI. That is, male teachers, teachers with lower qualification levels and the youngest and oldest teachers show lower participation rates in PDI (OECD 2010). The OECD (2010) underlines that the variability in teaching quality is only explained to a limited degree by characteristics such as formal education, teachers' personality and experience. In the study of Wasik and Hindman (2011) (preschool) teachers' formal education and experience did not play an important role in their practices or in student outcomes. Saka (2013) investigated the effects of demographic factors (i.e. gender, grade level and age) on teachers' participation in particular PDI. Significant differences were only found in respect to gender, as female teachers were more highly represented in a particular PDI. In addition to teachers' characteristics, Pehmer, Gröschner, and Seidel (2015) point to the importance of considering students' different preconditions in evaluating the effectiveness of PDI. In their research, the effects of the video-based PDI varied according to students' domain-specific self-concept. However, only very few studies (4%) investigate these specific teacher and student characteristics in relationship with the overall effects of the studied PDI.

Extended evaluative framework

The first research aim of this study focused on constructing an extended framework for the evaluation of the outcomes of a professional development initiative. As a result of a theoretical exploration and an in-depth analysis of the investigated outcome measures in our selected studies meeting our quality criteria (see method section), each of the components in Desimone's model (2009) could be further unravelled. This extended framework is presented in Figure 2.

The extended evaluative framework (Figure 2) presents more detailed information on elements that have to be taken into account when organising PDI and possible outcomes that can be aspired to

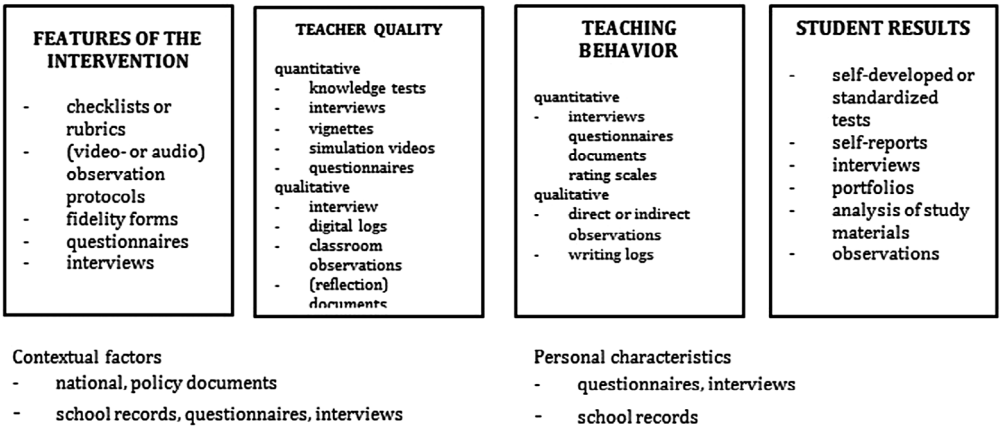

Figure 3. Overview of measures used for each component within the extended evaluative model for professional development.

Table 2. Overview of the main advantages and disadvantages of the measurements methods used in the selected studies.

	Advantage	Disadvantage
Quantitative measures (e.g. questionnaires; surveys)	Cost-effective	Response subject to bias
	Straightforward data gathering	Self-report
	Efficient in large samples	Hastily completion
	Gives evidence of associations	No accurate reflection of real practice
Qualitative measures (e.g. observations; interviews)	Real-time measures	Subject to bias
	Direct measures	Intense data gathering
	Rich data capturing	Intensive scoring and analysis
	Deep insights in complex interactions	Privacy and confidentiality issues
		Invasive method

and assessed after participation in a specific PDI. The appendix provides more detailed information on the outcome measures for each study included in the review.

Methods and measurement instruments

The extended evaluative model has been used to meet our second research aim, namely providing a consistent set of methods and measures to evaluate a PDI's effectiveness. The following sections provide a mainly descriptive overview of these currently used methods and measurement instruments to evaluate each (sub)component within the extended professional development framework. These are summarised in Figure 3. After this overview, we point additional attention to an evaluation of the advantages and disadvantages of the enumerated measures (see 'Overview of measurement instruments according to the evaluation framework' and Table 2). Further, some important preconditions in using these measures are mentioned.

Key features of PDI

Our analysis points to the importance of considering the evaluation of a professional development initiative's key features before, during and after the implementation of the PDI. *Before* the PDI, Hill et al. (2013) propose that more rigorous comparisons of PDI features should be executed at the initial stage of professional development. In our analysis, we found studies considering effective PDI features in the theoretical or method section of their manuscript. Here, the characteristics of the implemented initiative are mirrored to empirically investigate effective PDI characteristics. For instance, researchers stress the incorporation of a content and long-term focus (Morrison 2014), active participation

(Van den Bergh, Ros, and Beijaard 2015) and follow-up activities (Walker et al. 2012) in their professional development initiative to underpin its quality.

During the professional development initiative, fidelity to the PDI features can be assessed, that is the degree to which the PDI is delivered as intended (Ermeling 2010; Garet et al. 2008). In this respect, observers in the study of Garet et al. (2008) completed close-ended fidelity forms during institutes and seminars to assure fidelity to the PDI. Wasik and Hindman (2011) assess fidelity by completing checklists during teacher observations. They found a link between variation in fidelity and child outcomes, illuminating which particular aspects of the PDI were linked to children's gains (e.g. explicit instruction in PCK; immediate individualised feedback). Doppelt et al. (2009) coded workshops by means of video analysis to verify whether workshops were provided productively (e.g. presence of PCK; teacher reflections).

The features of PDI can also be considered *afterwards*. A minority of the studies, those mostly applying a quasi-experimental design, explicitly focus on a comparison of features after an intervention to design future PDI. For instance, Zhang et al. (2011) compare three types of videos (i.e. published, teacher' own and colleagues' videos) in a teacher's PDI to describe each type of affordances and challenges. Kiemer et al. (2015) found more positive effects of a video-based PDI compared to a traditional PDI to promote classroom discourse. Their video-based PDI seemed to promote a stronger learner community (cf. collaboration) and examples from their own practice (cf. ownership). Others query specific PDI features in survey research. For instance, Buczynski and Hansen (2010) focused on core features of the PDI (e.g. satisfaction with content delivery, instructional strategies and/or future PDI needs). Ingvarson, Meiers, and Beavis (2005) evaluated more structural features of the PDI (e.g. contact hour, time span, sufficient time, collective and/or participation) with self-developed instruments. To conclude, effective PDI features can be evaluated before, during or after a PDI, and have mostly been assessed by means of a theoretical consideration, checklists, fidelity protocols or surveys.

Increased teacher quality

In the selected studies, teacher quality has been assessed in a variety of ways. As to teachers' *cognitive goals*, researchers have mostly applied self-developed tests or questionnaires (e.g. Buczynski and Hansen 2010; Fishman et al. 2003; Goldschmidt and Phelps 2010; McCutchen et al. 2002) measuring teachers' pedagogical (content) knowledge. Here, mostly pre- and post-tests are used to map teachers' (knowledge) gains scores. Some researchers used interviews (e.g. to measure conceptual understandings concerning self-regulated learning (Butler et al. 2004) or citizenship education (Willemse et al. 2015)). To assess teachers' improved *skills*, video tests and vignettes were mainly used wherein classroom situations are shown or simulated. Vogt and Rogalla (2009) used video tests wherein teachers had to instantly decide to stop a video when perceiving non-adaptive situations, express their perceptions and suggest a more adaptive alternative to the teachers' action. Dedousis-Wallace et al. (2014) used vignettes, wherein teachers where provided with different bullying scenarios to measure teachers' self-predicted responses to these situation. As to teachers' *affective goals* (i.e. attitudes and beliefs), teachers' interviews (James and McCormick 2009), digital writing logs (Bakkenes, Vermunt, and Wubbels 2010) or surveys (Saka 2013) have been used. Classroom observations can also be used to assess general teacher quality. For instance, Domitrovich et al. (2009) applied the classroom assessment scoring system (CLASS), and observational measures that assess 10 dimensions of teaching quality, identified through a systematic literature review. In sum, teacher quality has mainly been evaluated by means of (knowledge) tests, questionnaires or interviews.

Change in instruction

Changes in instruction have overall been measured with direct or indirect observations, either immediately in classrooms (e.g. Buczynski and Hansen 2010; Franke et al. 2001; Lee et al. 2004) or afterwards by coding videos from classroom situations or workshops (e.g. Chamberlin 2005; Doppelt et al. 2009). Garet et al. (2008) provide information on how the development of a classroom observation protocol is conducted and how such a protocol can be constructed. Furthermore, the analysis of interviews

(Butler et al. 2004), digital writing logs (Bakkenes, Vermunt, and Wubbels 2010), questionnaires or surveys (Garet et al. 2001; James and McCormick 2009) and documents (e.g. observation notes; learner materials (Chamberlin 2005)) has been used to assess changes in instructional strategies or practices. To examine interaction patterns, Domitrovich et al. (2009) applied the teaching-style rating scale to identity different interaction patterns.

Improved student learning

Considering students' increased performances, data can be gathered directly from students. In this respect, researchers have applied self-developed (Kiemer et al. 2015; Saxe, Gearhart, and Nasir 2001; Vogt and Rogalla 2009) or more standardised tests (Buczynski and Hansen 2010; McCutchen et al. 2002). Fishman et al. (2003), however, advise against relying too heavily on distal measures, such as standardised tests, as they do not reference directly to the implemented curriculum at that moment. Data on student learning can also be gathered indirectly by questioning teachers or parents on their perceived student learning by means of interviews or rating scales (Bierman et al. 2008; Butler et al. 2004). Wallace (2009) states that gains scores (i.e. students learning gains over an extended period of time) have been successfully employed in research as a barometer of growth, even though their use has been criticised because of the main influences besides teaching that can affect student scores. It is notable that the studies included in our review mostly apply product-oriented measures (i.e. gathered after task execution by means of tests) and very little process-oriented measures are used to evaluate students' learning processes during task execution (e.g. observations (Bierman et al. 2008)).

School context and personal characteristics

Contextual factors, such as school context and personal teacher and student characteristics, have mainly been mapped by means of questionnaires. For example, Hofman and Dijkstra (2010) used a questionnaire to query teachers' motives for participating in PDI. James and McCormick (2009) questioned the dimension of school management practices (e.g. involving staff in decision-making; supporting professional development) with a questionnaire. Information on teachers' perceptions of principal support (Supovitz and Turner 2000) and students' gender and SES (Antoniou and Kyriakides 2013) was gathered by means of a questionnaire. Interviews can also be used to map contextual or school organisational conditions (e.g. coordination of the PD across school personnel (Butler et al. 2004)). Here, information was requested from teachers directly or indirectly through the school principal.

Overview of measurement instruments according to the evaluation framework

Figure 3 represents an overview of the main applied measurement instruments according to each component in the PD framework.

When overviewing and evaluating the selected articles on their applied measurement instruments, two general remarks can be made. First, some authors have explicitly addressed the advantages and disadvantages of the applied measurement instruments in their study (see Table 2 for an overview). We believe these (dis)advantages should be carefully considered when planning to evaluate the effectiveness of a professional development initiative. For instance, quantitative measures (e.g. questionnaires; surveys) are more cost-effective than intensive data gathering techniques (e.g. observations; interviews). Data gathering is straightforward and efficient in large samples. Furthermore, quantitative data gives evidence of associations (James and McCormick 2009). However, the responses might be subject to bias. For instance, it is unclear whether teachers can adequately report on the school and curricular practice (Supovitz and Turner 2000), and whether students correctly understand questions (Mikami et al. 2011). Rather long questionnaires might be hastily completed, influencing the measures' reliability (Dedousis-Wallace et al. 2014), and self-reports might not be a completely accurate reflection of the real practice. Vignettes might allow participants to better contextualise the questions, and structurally formulate answers in their own words (Dedousis-Wallace et al. 2014; Vogt and Rogalla 2009). Vogt and Rogalla (2009), for example, used video tests to measures teachers' skills as an alternative to real-time observation. In this way, all participants received the same (standardised) test and their answers could

be registered immediately. Field notes might be subject to bias as they cannot reflect every spoken word and depend upon the recorders' accuracy to capture the addressed topics (Levine and Marcus 2010). Real-time classroom of video observations is advantageous on the one hand, as they provide direct measures of teachers' changes in instruction, offer rich data to capture complex interactions (Bakkenes, Vermunt, and Wubbels 2010) and might give more insights into possible explanations for particular findings (James and McCormick 2009). However, they mostly require intensive scoring and analysis, and privacy and confidentiality issues must be addressed (Bakkenes, Vermunt, and Wubbels 2010). Using (video) observations can be an invasive method that only a small number of teachers might be willing to undergo as they might feel anxious or stressed when being observed (Kiemer et al. 2015; Liu 2013) or behave differently simply because they are being observed (Liang 2015; Merrett 2006).

Second, authors have underlined the importance of using multiple measures and informants for different evaluation components to map the effectiveness of PDI. In view of the importance of this data triangulation, it was encouraging to find that many researchers already draw on both quantitative and qualitative results from multiple data sources (e.g. Abuhmaid 2011; Dedousis-Wallace et al. 2014; Kiemer et al. 2015; Morrison 2014; Wasik and Hindman 2011). By applying these diverse measures, some authors were able to detect contrasting findings (Lee et al. 2004; Mikami et al. 2011). For example, Lee et al. (2004) found a discrepancy between teachers' perceptions of their improved science knowledge and practice (as indicated in questionnaires and interviews) and the lack of significant change in their actual practices (as evidenced in the classroom observations). They ascribe this to the fact that effective practices might require more than strong beliefs, but need deep and robust knowledge of subject matter content and content-specific teaching strategies. From classroom observations, Mikami et al. (2011) observed positive effects from a PDI on students' peer relationships. This positive peer interaction pattern, however, was not shown in students' self-report data. By combining different data sources, these discrepancies were identified and possible intermediating factors could be identified. Further, Muijs and Lindsay (2008) found a relationship between levels of evaluation (cf. Guskey's level of evaluation, see Table 1) employed and evaluation methods used. More specifically, schools that appear to evaluate (continuous) professional development at more levels appear more sophisticated in the use of multiple research methods. Additionally, in studies included in our analysis in which relationships were considered between PDI and several teacher and student outcomes, a more varied pallet of both qualitative and quantitative outcome measures was applied (e.g. Buczynski and Hansen 2010; Wasik and Hindman 2011).

Conclusion and discussion

Current study

This study focused on two specific research aims: (1) providing an extended evaluative framework for PDI and (2) overviewing currently used methods and measurement instruments to measure these PDI outcomes. By means of a systematic narrative synthesis whereby studies were selected based on five important quality criteria (see method section), Desimone's (2009) model of professional development was extended with diverse subcomponents to be investigated in PDI evaluation (Figure 1). For example, Desimone's (2009) five features of the intervention were extended to nine crucial features. In addition, 'teacher quality', 'teacher instruction', 'learning outcomes', 'contextual factors' and 'personal charactistics' were further extended. In this respect, this study adds to the literature as, to date, no such extended framework was available to more directly guide researchers and practitioners in the evaluation of PDI. Furthermore, this study addresses an important limitation in previous literature as it unites views on important components in PDI evaluation (Desimone 2009) and evaluation models and methods of professional development (Guskey 2000, 2014). As to the second research aim, each component is linked with possible measurement instruments. It is acknowledged that evaluators of PDI might not be able to evaluate each one of the (sub)components presented in the framework as this requires extensive resources. However, the framework provides an overall view on potentially

important components and guides evaluators to purposefully decide on the (sub)components to be investigated. It informs them as to where possible unintended consequences might turn up and links this with other components in the extended framework. Further, this model wants to stress the intertwining of components and recursive, interactive relationships between the components (Desimone 2009; Levenson and Gal 2013) and the cyclical nature of PDI evaluation (King 2014; Opfer and Pedder 2011), in contrast with Guskey's (2000, 2014) hierarchical levels of evaluation.

Limitations and implications for research and practice

To conclude, some limitations of the study are addressed and presented together with implications for educational researchers and developers of professional development.

For educational researchers, it could be noticed that only few researchers theoretically frame their research within an overall evaluative framework. Therefore, a first recommendation is to use the extended evaluative framework proposed in this study in future empirical work (e.g. research articles, dissertations and/or educational reports) for theoretically introducing the evaluation of PDI, reporting on results and discussing the effectiveness of the specific PDI under investigation. More particularly, research questions could be operationalised according to the specified subcategories of the five general components in the framework (Figure 2). When using this common theoretical base, comparability between different studies will be facilitated. Second, in line with Ingvarson, Meiers, and Beavis (2005) and Desimone (2009), we underline the importance of documenting more extensively on used methodology and measurement instruments. Many articles lack the inclusion of complete measurement instruments or coding schemes in their appendices. Therefore, educational researchers are asked to include transparent descriptions of data collection, analysis and measurement instruments used and reliability of the measurement instruments when documenting PDI research. This would provide educational researchers and developers of PDI user-friendly and evidence-based material to be implemented in future PDI (research) across different designs. More explicit reflections on the strengths and weaknesses of the applied measurement instruments should be provided. In this way, others can informatively decide on the appropriateness of including these specific instruments into their own PDI and/or research. This could lead to an ongoing instrument refinement across a variety of classroom settings. Third, educational researchers are encouraged to continue to unravel the (sub) components included in the extended framework. Although a search strategy was applied to yield as many studies as possible, particular studies might have slipped through the net and might therefore not be included into our review. As already mentioned, more research is warranted concerning several elements integrated into the extended framework. For example, the role of professional identity in teachers' professional development (Beijaard, Meijer, and Verloop 2004). Although specific components of teachers' professional identity have been identified in our analysis (e.g. teachers' self-efficacy as part of quality of the teacher [Morrison 2014;]), no results were yielded concerning impact studies investigating influences of PDI on teachers' professional identity as a whole. Future research could also investigate in greater depth the role of contextual factors (e.g. principal supportiveness [Supovitz and Turner 2000]; commitment of the school administration to the PDI [Voerman et al. 2015]), teacher characteristics (e.g. teachers work load or incentives to undertake PDI [Abuhmaid 2011]) and student characteristics (e.g. self-concept [Pehmer, Gröschner, and Seidel 2015]) that might influence the effectiveness of PDI. In line with Antoniou and Kyriakides (2013), stronger links in future research should be made between PDI research and educational effectiveness research (ERR) to establish an effective approach to teacher PDI and to further unravel the relationship between effective PDI features, changes in teacher quality, instruction and student outcomes. Furthermore, the framework is based on PDI in formal education. Future research might also include PDI for other target groups such as teachers in adult education and university professors (Roblin and Margalef 2013) or focus on new forms of professional development such as technology-related and online professional development (Smith and Sivo 2012; Walker et al. 2012). Research into these new forms of PDI might possibly yield additional subcomponents for investigation.

As to the developers (and evaluators) of PDI, this extended framework represents a starting point when developing and evaluating PDI. In this respect, Guskey (2000, 2014) proposes a 'planning backwards' approach which starts with identifying efforts that have produced demonstrable evidence of success. Inspired by his suggestion and our own research findings, we suggest a seven-step approach to a focused and systematic evaluation of PDI.

Prior the start of the PDI, three major steps should be taken into account. First, to set up realistic expectations, it is important to consider in which way the PDI meets the effective PDI features as specified in Figure 1 (step 1). In a second step, the extended framework can be used as a starting point to identify the aspired outcomes on various subcomponents in teacher quality, change in instruction and student outcomes. It is important to define clear and measurable objectives (Linn et al. 2010) for both expected and unexpected outcomes. It is also important to consider what might impede the implementation of certain aspects in the PD as this might help the planning, offering and follow-up of the intervention (Morrison 2014) (step 2). In a third step, both quantitative and qualitative research instruments can be identified for each (sub)component with a view to integrating the results obtained with these different measures into a holistic picture (James and McCormick 2009) (step 3). In search of appropriate measures, it might be interesting for designers of professional development to conduct a pilot study to check clarity, instruction and lay out of the measurement instruments (Abuhmaid 2011). For instance, Abuhmaid (2011) found that open-ended items were generally left unanswered, which led to reformatting the open-ended items to become clearly structured, easier and faster to complete (or items requiring rewording).

During the intervention, a fourth step in the focused and systematic evaluation is data collection on the outcome measures specified in step 2 with the instruments identified in step 3. Additionally, a feasibility study can be conducted to verify whether changes in outcome measures are due to critical PDI features (Pehmer, Gröschner, and Seidel 2015) (step 4).

After the PDI, data collection can continue by gathering data shortly after the PDI on the one hand and after a longer period of time on the other hand (step 5). In this respect, an important aspect in the evaluation of PDI is its sustainability in time and the generativity of the PDI. The importance of a long-term evaluation is a research challenge addressed by many researchers (Avalos 2011; Franke et al. 2001; Lee et al. 2004; Morge, Toczek, and Chakroun 2010; Muijs and Lindsay 2008; Schachter 2015). For instance, Antoniou and Kyriakides (2013) did not find any improvement or decline of the PDI after one year. In contrast, Franke et al. (2001) did find some effects four years after the professional development ended. A long-term evaluation makes it possible to more surely conclude that change actually took place (Morge, Toczek, and Chakroun 2010), and that the PDI persists in the complex day-to-day practice after the PDI is finished (Ponte et al. 2004). Further, some effects might only be able to be traced after longer periods of time, such as effects on pupils (Muijs and Lindsay 2008). A sixth step involves data analysis and interpretation of the results. Here, it is important to reconsider all effective PDI features and aspired outcomes defined in steps 1 and 2 (step 6). In a seventh and final step, a general PDI evaluation can be made, reconsidering the overall effects in relation to the effective PDI features and used (the advantages and disadvantages) measurement instruments (step 7).

To undertake this evaluation in a focused and systematic way, collaboration and partnerships with (social) scientists in setting up and evaluating professional development might be highly advisable, for instance, for transferring specific content knowledge or teaching practices to teachers (e.g. Morrison 2014; Willemse et al. 2015) and also to conduct the more intensive and complex data analysis techniques in step 6 regarding psychometric testing and scaling, research design, analysis and information dissemination (Avalos 2011; Stake 1967). It might also help (a) to implement more quasi-experimental designs to compare different PDI and develop the most effective one in a particular context (e.g. McCutchen et al. 2002; Saxe, Gearhart, and Nasir 2001; Van Keer and Verhaeghe 2005), (b) to compare multiple data-sets, indicating consistency across contexts (Wallace 2009) and (c) to narrow the research–practice gap where teachers have more access to and understanding of empirical research findings on the evaluation of PDI (Earley and Porritt 2014).

To conclude, this study addressed an important conceptual and methodological challenge in the evaluation of PDI. More particularly, an extended evaluative framework is proposed for the focused and systematic evaluation of PDI by mapping the outcomes that can be evaluated (focused) by means of particular measurement instruments (systematic). The extended framework and the seven-step approach for the evaluation of PDI can be used by developers and evaluators of PDI and educational researchers.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Emmelien Merchie, PhD, is a postdoctoral assistant at the Department of Educational Studies, Ghent University (Belgium). Her research interests are in the field of stimulating text-learning strategy use and knowledge acquisition in elementary education and professional development in general.

Melissa Tuytens, PhD, is a postdoctoral researcher affiliated with the Department of Educational Studies, Ghent University (Belgium). Her research interests are situated in the field of personnel policy within schools, school leadership and school policy in general.

Geert Devos, PhD, is an associate professor in educational administration at the Department of Educational Studies, Ghent University (Belgium). His current research areas include school leadership, educational policy and school improvement.

Ruben Vanderlinde, PhD, is a tenure track professor at the Department of Educational Studies, Ghent University (Belgium). His research interests are in the field of educational innovation, teacher training and professionalization, and the integration of Information and Communication Technologies (ICT) in education.

References

Abuhmaid, A. 2011. "Ict Training Courses for Teacher Professional Development in Jordan." *Turkish Online Journal of Educational Technology* 10 (4): 195–210.

Antoniou, P., and L. Kyriakides. 2013. "A Dynamic Integrated Approach to Teacher Professional Development: Impact and Sustainability of the Effects on Improving Teacher Behaviour and Student Outcomes." *Teaching and Teacher Education* 29 (1): 1–12. doi:10.1016/j.tate.2012.08.001.

Avalos, B. 2011. "Teacher Professional Development in Teaching and Teacher Education over Ten Years." *Teaching and Teacher Education* 27 (1): 10–20. doi:10.1016/j.tate.2010.08.007.

Baeten, M., and M. Simons. 2014. "Student Teachers' Team Teaching: Models, Effects, and Conditions for Implementation." *Teaching and Teacher Education* 41: 92–110. doi:10.1016/j.tate.2014.03.010.

Bakkenes, I., J. D. Vermunt, and T. Wubbels. 2010. "Teacher Learning in the Context of Educational Innovation: Learning Activities and Learning Outcomes of Experienced Teachers." *Learning and Instruction* 20 (6): 533–548. doi:10.1016/j.learninstruc.2009.09.001.

Beijaard, D., P. C. Meijer, and N. Verloop. 2004. "Reconsidering research on teachers' professional identity." *Teaching and Teacher Education* 20 (2): 107–128. doi:10.1016/j.tate.2003.07.001.

Bierman, K. L., C. E. Domitrovich, R. L. Nix, S. D. Gest, J. A. Welsh, M. T. Greenberg, and C. Blair. 2008. "Promoting Academic and Social-emotional School Readiness: The Head Start REDI Program." *Child Development* 79 (6): 1802–1817. doi:10.1111/j.1467-8624.2008.01227.x.

Borko, H. 2004. "Professional Development and Teacher Learning: Mapping the Terrain." *Educational Researcher* 33 (8): 3–15. doi:10.3102/0013189X033008003.

Boston, M. D. 2013. "Connecting Changes in Secondary Mathematics Teachers' Knowledge to Their Experiences in a Professional Development Workshop." *Journal of Mathematics Teacher Education* 16 (1): 7–31. doi:10.1007/s10857-012-9211-6.

Buczynski, S., and C. B. Hansen. 2010. "Impact of Professional Development on Teacher Practice: Uncovering Connections." *Teaching and Teacher Education* 26 (3): 599–607. doi:10.1016/j.tate.2009.09.006.

Butler, D. L., H. N. Lauscher, S. Jarvis-Selinger, and B. Beckingham. 2004. "Collaboration and Self-regulation in Teachers' Professional Development." *Teaching and Teacher Education* 20 (5): 435–455. doi:10.1016/j.tate.2004.04.003.

Chamberlin, M. T. 2005. "Teachers' Discussions of Students' Thinking: Meeting the Challenge of Attending to Students' Thinking." *Journal of Mathematics Teacher Education* 8 (2): 141–170. doi:10.1007/s10857-005-4770-4.

Cheng, M. M. H., and W. W. M. So. 2012. "Analysing Teacher Professional Development through Professional Dialogue: An Investigation into a University-School Partnership Project on Enquiry Learning." *Journal of Education for Teaching* 38 (3): 323–341. doi:10.1080/02607476.2012.668331.

Cohen, D. K., and H. C. Hill. 2000. "Instructional Policy and Classroom Performance: The Mathematics Reform in California." *Teachers College Record* 102 (2): 294–343. doi:10.1111/0161-4681.00057.

Consuegra, E. 2015. *Gendered Teacher–Student Classroom Interactions in Secondary Education: Perception, Reality and Professionalism*, 171–200. Brussels: VUB.

De Neve, D., G. Devos, and M. Tuytens. 2015. "The Importance of Job Resources and Self-efficacy for Beginning Teachers' Professional Learning in Differentiated Instruction." *Teaching and Teacher Education* 47: 30–41. doi:10.1016/j.tate.2014.12.003.

De Vries, S., W. J. C. M. van de Grift, and E. P. W. A. Jansen. 2014. "How Teachers' Beliefs about Learning and Teaching Relate to Their Continuing Professional Development." *Teachers and Teaching* 20 (3): 338–357.

Dedousis-Wallace, A., R. Shute, M. Varlow, R. Murrihy, and T. Kidman. 2014. "Predictors of Teacher Intervention in Indirect Bullying at School and Outcome of a Professional Development Presentation for Teachers." *Educational Psychology*, 34 (7): 862–875. doi:10.1080/01443410.2013.785385.

Desimone, L. M. 2009. "Improving Impact Studies of Teachers' Professional Development: Toward Better Conceptualizations and Measures." *Educational Researcher* 38 (3): 181–199. doi:10.3102/0013189X0833114.

Desimone, L. M., A. C. Porter, M. S. Garet, K. S. Yoon, and B. F. Birman. 2002. "Effects of Professional Development on Teachers' Instruction: Results from a Three-year Longitudinal Study." *Educational Evaluation and Policy Analysis* 24 (2): 81–112. doi:10.3102/01623737024002081.

Dochy, F., M. Segers, and M. M. Buehl. 1999. "The Relationship between Assessment Practices and Outcomes of Studies: The Case of Research on Prior Knowledge." *Review of Educational Research1* 69 (2): 145–186. doi:10.3102/00346543069002145.

Domitrovich, C. E., S. D. Gest, S. Gill, K. L. Bierman, J. A. Welsh, and D. Jones. 2009. "Fostering High-Quality Teaching with an Enriched Curriculum and Professional Development Support: The Head Start REDI Program." *American Educational Research Journal* 46 (2): 567–597. doi:10.3102/0002831208328089.

Doppelt, Y., C. D. Schunn, E. M. Silk, M. M. Mehalik, B. Reynolds, and E. Ward. 2009. "Evaluating the Impact of a Facilitated Learning Community Approach to Professional Development on Teacher Practice and Student Achievement." *Research in Science & Technological Education* 27 (3): 339–354. doi:10.1080/02635140903166026.

Earley, P., and V. Porritt. 2014. "Evaluating the Impact of Professional Development: The Need for a Student-Focused Approach." *Professional Development in Education* 40 (1): 112–129. doi:10.1080/19415257.2013.798741.

Ekanayake, S. Y., and J. Wishart. 2014. "Integrating Mobile Phones into Teaching and Learning: A Case Study of Teacher Training through Professional Development Workshops." *British Journal of Educational Technology* 46 (1): 173–189. doi:10.1111/bjet.12131.

Ermeling, B. A. 2010. "Tracing the Effects of Teacher Inquiry on Classroom Practice." *Teaching and Teacher Education* 26 (3): 377–388. doi:10.1016/j.tate.2009.02.019.

Fishman, B. J., R. W. Marx, S. Best, and R. T. Tal. 2003. "Linking Teacher and Student Learning to Improve Professional Development in Systemic Reform." *Teaching and Teacher Education* 19 (6): 643–658. doi:10.1016/S0742-051X(03)00059-3.

Franke, M. L., T. P. Carpenter, L. Levi, and E. Fennema. 2001. "Capturing Teachers' Generative Change: A Follow-up Study of Professional Development in Mathematics." *American Educational Research Journal* 38 (3): 653–689. doi:10.3102/00028312038003653.

Garet, M. S., S. Cronen, M. Eaton, and A. Kurki. 2008. *The Impact of Two Professional Development Interventions on Early Reading Instruction and Achievement. NCEE 2008-4030. National Center for ….* http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:The+Impact+of+Two+Professional+Development+Interventions+on+Early+Reading+Instruction+and+Achievement#0\nhttp://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:The+Impact+of+Two+Professi.

Garet, M. S., A. C. Porter, L. Desimone, B. F. Birman, and K. S. Yoon. 2001. "What Makes Professional Development Effective? Results from a National Sample of Teachers." *American Educational Research Journal* 38 (4): 915–945. doi:10.3102/00028312038004915.

Goldschmidt, P., and G. Phelps. 2010. "Does Teacher Professional Development Affect Content and Pedagogical Knowledge: How Much and for How Long?" *Economics of Education Review* 29 (3): 432–439. doi:10.1016/j.econedurev.2009.10.002.

Grosemans, I., A. Boon, C. Verclairen, F. Dochy, and E. I. Kyndt. 2015. "Informal Learning of Primary School Teachers: Considering the Role of Teaching Experience and School Culture." *Teaching and Teacher Education* 47: 151–161.

Guskey, T. R. 2000. *Evaluating Professional Development*. Thousand Oaks, CA: Corwin press.

Guskey, T. R. 2003. "Analyzing Lists of the Characteristics of Effective Professional Development to Promote Visionary Leadership." *NASSP Bulletin* 87 (637): 4–20. doi:10.1177/019263650308763702.

Guskey, T. R. 2014. "Evaluating Professional Learning." In *International Handbook of Research in Professional and Practice-based Learning*, edited by S. Billett, 1215–1235. New York, NY: Springer International.

Hammerness, K., L. Darling-Hammond, J. Bransford, D. Berliner, M. Cochran-Smith, M. McDonald, and K. Zeichner. 2005. "How Teachers Learn and Develop." In *Preparing Teachers for a Changing World What Teachers Should Learn and Be Able to Do*, 358–389. https://login.proxy.library.msstate.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2005-13868-010&login.asp&site=ehost-live.

Hammond, R. L. 1973. *Evaluation at the Local Level*. Arizona: U.S. Department of Health, Education & Welfare, Office of Education.

Hattie, J. 2009. *Visible Learning: A Synthesis of over 800 Meta-analyses Relating to Achievement*. New York: Routledge.

Hill, H. C., M. Beisiegel, and R. Jacob. 2013. "Professional development research: Concensus, crossroads, and challenges." *Educational Researcher* 42: 476–487. doi:10.3102/0013189X13512674.

Hofman, R. H., and B. J. Dijkstra. 2010. "Effective Teacher Professionalization in Networks?" *Teaching and Teacher Education* 26 (4): 1031–1040. doi:10.1016/j.tate.2009.10.046.

Ingvarson, L., M. Meiers, and A. Beavis. 2005. "Factors Affecting the Impact of Professional Development Programs on Teachers' Knowledge, Practice, Student Outcomes & Efficacy." *Education Policy Analysis Archives* 13 (10): 1–28. doi:10.14507/epaa.v13n10.2005.

James, M., and R. McCormick. 2009. "Teachers Learning How to Learn." *Teaching and Teacher Education* 25: 973–982. doi:10.1016/j.tate.2009.02.023.

Kang, H. S., J. Cha, and B.-W. Ha. 2013. "What Should We Consider in Teachers' Professional Development Impact Studies? Based on the Conceptual Framework of Desimone." *Creative Education* 4: 11–18. doi:10.4236/ce.2013.44A003.

Kazemi, E., and M. L. Franke. 2004. "Teacher Learning in Mathematics: Using Student Work to Promote Collective Inquiry." *Journal of Mathematics Teacher Education* 7: 203–235. doi:10.1023/B:JMTE.0000033084.26326.19.

Kedzior, M. M. 2004. "Teacher Professional Development." *Education Policy Brief* 15: 1–6.

Kiemer, K., A. Gröschner, A. Pehmer, and T. Seidel. 2015. "Effects of a Classroom Discourse Intervention on Teachers' Practice and Students' Motivation to Learn Mathematics and Science." *Learning and Instruction* 35: 94–103. doi:10.1016/j.learninstruc.2014.10.003.

King, F. 2014. "Evaluating the Impact of Teacher Professional Development: An Evidence-based Framework." *Professional Development in Education* 40 (1): 89–111.

Kirkpatrick, D. L. 1994. *Evaluating Training Programs: The Four Levels*. San Francisco, CA: Berrett-Koehler.

Kirkevold, M. 1997. "lntegrative nursing research-an important strategy to further the development of nursing science and Practice." *Journal of Advanced Nursing* 25 (5): 977–984. doi:10.1046/j.1365-2648.1997.1997025977.x.

Lee, O., J. E. Hart, P. Cuevas, and C. Enders. 2004. "Professional Development in Inquiry-based Science for Elementary Teachers of Diverse Student Groups." *Journal of Research in Science Teaching* 41 (10): 1021–1043. doi:10.1002/tea.20037.

Levenson, E., and H. Gal. 2013. "Insights From a Teacher Professional Development Course: Rona's Changing Perspectives Regarding Mathematically-talented Students." *International Journal of Science and Mathematics Education* 11: 1087–1114.

Levine, T. H., and A. S. Marcus. 2010. "How the Structure and Focus of Teachers' Collaborative Activities Facilitate and Constrain Teacher Learning." *Teaching and Teacher Education* 26 (3): 389–398. doi:10.1016/j.tate.2009.03.001.

Liang, J. 2015. "Live Video Classroom Observation: An Effective Approach to Reducing Reactivity in Collecting Observational Information for Teacher Professional Development." *Journal of Education for Teaching* 41 (3): 235–253. doi:10.1080/02607476.2015.1045314.

Linn, G. B., P. Gill, R. Sherman, V. Vaughn, and J. Mixon. 2010. "Evaluating the long-term impact of professional development." *Professional Development in Education* 36 (4): 679–682. doi:10.1080/19415250903109288.

Liu, S.-H. 2013. "Teacher Professional Development for Technology Integration in a Primary School Learning Community." *Technology, Pedagogy and Education* 22 (1): 37–54. doi:10.1080/1475939X.2012.719398.

Loewenberg Ball, D., M. H. Thames, and G. Phelps. 2008. "Content Knowledge for Teaching: What Makes It Special?" *Journal of Teacher Education* 59 (5): 389–407. doi:10.1177/0022487108324554.

McCutchen, D., R. D. Abbott, L. B. Green, S. N. Beretvas, S. Cox, N. S. Potter, and T. Quiroga. 2002. "Beginning Literacy: Links among Teacher Knowledge, Teacher Practice, and Student Learning." *Journal of Learning Disabilities* 35: 69–86.

Merrett, F. 2006. "Reflections on the Hawthorne Effect." *Educational Psychology* 26 (1): 143–146. doi:10.1080/01443410500341080.

Mikami, A. Y., A. Gregory, J. P. Allen, R. C. Pianta, and J. Lun. 2011. "Effects of a Teacher Professional Development Intervention on Peer Relationships in Secondary Classrooms." *School Psychology Review* 40 (3): 367–385. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3379816&tool=pmcentrez&rendertype=abstract.

Mishra, P., and M. J. Koehler. 2006. "Technological Pedagogical Content Knowledge: A Framework for Teacher Knowledge." *Teachers College Record* 180 (6): 1017–1054.

Morge, L., M. C. Toczek, and N. Chakroun. 2010. "A Training Programme on Managing Science Class Interactions: Its Impact on Teachers' Practises and on Their Pupils' Achievement." *Teaching and Teacher Education* 26 (3): 415–426. doi:10.1016/j.tate.2009.05.008.

Morrison, J. A. 2014. "Scientists' Participation in Teacher Professional Development: The Impact on Fourth to Eighth Grade Teachers' Understanding and Implementation of Inquiry Science." *International Journal of Science and Mathematics Education* 12: 793–816.

Muijs, D., and G. Lindsay. 2008. "Where Are We at? An Empirical Study of Levels and Methods of Evaluating Continuing Professional Development." *British Educational Research Journal* 34 (2): 195–211. doi:10.1080/01411920701532194.
National Research Council. 2010. *Preparing Teachers: Building Evidence for Sound Policy*. Washington, DC: Committee on the Study of Teacher Preparation Programs in the United States. Division of Behavioral and Social Science and Education.
Niess, M. L. 2005. "Preparing Teachers to Teach Science and Mathematics with Technology: Developing a Technology Pedagogical Content Knowledge." *Teaching and Teacher Education* 21 (5): 509–523. doi:10.1016/j.tate.2005.03.006.
Nir, A. E., and R. Bogler. 2008. "The Antecedents of Teacher Satisfaction with Professional Development Programs." *Teaching and Teacher Education* 24 (2): 377–386. doi:10.1016/j.tate.2007.03.002.
Norton, A. H., and A. McCloskey. 2008. "Teaching Experiments and Professional Development." *Journal of Mathematics Teacher Education* 11 (4): 285–305. doi:10.1007/s10857-008-9076-x.
OECD. 2010. *Teachers' Professional Development. Europe in International Comparison. An Analysis of Teachers' Professional Development Based on the OECD's Teaching and Learning International Survey (TALIS)*. Professor Jaap Scheerens, University of Twente. doi:10.2766/63494.
Opfer, V. D., and D. Pedder. 2011. "The Lost Promise of Teacher Professional Development in England." *European Journal of Teacher Education* 34 (1): 3–24.
Pehmer, A.-K., A. Gröschner, and T. Seidel. 2015. "How Teacher Professional Development regarding Classroom Dialogue Affects Students' Higher-order Learning." *Teaching and Teacher Education* 47: 108–119. doi:10.1016/j.tate.2014.12.007.
Ponte, P., J. Ax, D. Beijaard, and T. Wubbels. 2004. "Teachers' Development of Professional Knowledge through Action Research and the Facilitation of This by Teacher Educators." *Teaching and Teacher Education* 20 (6): 571–588. doi:10.1016/j.tate.2004.06.003.
Popay, J., H. Roberts, A. Sowden, M. Petticrew, L. Arai, M. Rodgers, N. Britten, K. Roen, and S. Duffy. 2006. *Guidance on the Conduct of Narrative Synthesis in Systematic Reviews*. A Product from the ESRC Methods Programme: Institute for Health Research, Lancaster University. doi:10.13140/2.1.1018.4643.
Postholm, M. B. 2012. "Teachers' Professional Development: A Theoretical Review." *Educational Research* 54 (4): 405–429. doi:10.1080/00131881.2012.734725.
Richter, D., M. Kunter, U. Klusmann, O. Lüdtke, and J. Baumert. 2011. "Professional Development across the Teaching Career: Teachers' Uptake of Formal and Informal Learning Opportunities." *Teaching and Teacher Education* 27 (1): 116–126. doi:10.1016/j.tate.2010.07.008.
Roblin, N. P., and L. Margalef. 2013. "Learning from Dilemmas: Teacher Professional Development through Collaborative Action and Reflection." *Teachers and Teaching* 19 (1): 18–32. doi:10.1080/13540602.2013.744196.
Saka, Y. 2013. "Who Are the Science Teachers That Seek Professional Development in Research Experience for Teachers (RET's)? Implications for Teacher Professional Development." *Journal of Science Education and Technology* 22: 934–951.
Saxe, G. B., M. Gearhart, and N. S. Nasir. 2001. "Enhancing Students' Understanding of Mathematics: A Study of Three Contrasting Approaches to Professional Support 1." *Journal of Mathematics Teacher Education* 4 (1): 55–79. doi:10.1023/A:1009935100676.
Schachter, R. E. 2015. "An Analytic Study of the Professional Development Research in Early Childhood Education." *Early Education and Development* 26 (8): 1057–1085. doi:10.1080/10409289.2015.100933.
Shulman, L. S. 1986. "Those Who Understand: Knowledge Growth in Teaching." *Educational Researcher* 15 (2): 4–14. doi:10.3102/0013189X015002004.
Smith, J. A., and S. A. Sivo. 2012. "Predicting Continued Use of Online Teacher Professional Development and the Influence of Social Presence and Sociability." *British Journal of Educational Technology* 43 (6): 871–882. doi:10.1111/j.1467-8535.2011.01223.x.
Stake, R. E. 1967. "The Countenance of Educational Evaluation." *Teachers College Record* 68 (7): 523–540.
Supovitz, J. A., and H. M. Turner. 2000. "The Effects of Professional Development on Science Teaching Practices and Classroom Culture." *Journal of Research in Science Teaching* 37 (9): 963–980. doi:10.1002/1098-2736(200011)37:9<963::AID-TEA6>3.0.CO;2-0.
Teemant, A., J. Wink, and S. Tyra. 2011. "Effects of Coaching on Teacher Use of Sociocultural Instructional Practices." *Teaching and Teacher Education* 27 (4): 683–693. doi:10.1016/j.tate.2010.11.006.
Tienken, C. H., and C. M. Achilles. 2003. "Changing Teacher Behavior and Improving Student Writing Achievement." *Planning and Changing* 34 (3&4): 153–168.
Timperley, H. 2008. "Teacher Professional Learning and Development." *The International Academy of Education* 1 (18): 1–30.
Timperley, H., A. Wilson, H. Barrar, and I. Fung. 2007. "Teacher Professional Learning and Development." *Best Evidence Synthesis Iteration. Education* 33 (8): 3–15. doi:10.1111/j.1744-7984.2007.00116.x.
Van den Bergh, L., A. Ros, and D. L. Beijaard. 2015. "Teacher Learning in the Context of a Continuing Professional Development Programme: A Case Study." *Teaching and Teacher Education* 47: 142–150. doi:10.1016/j.tate.2015.01.002.
Van Driel, J. H., and A. Berry. 2012. "Teacher Professional Development Focusing on Pedagogical Content Knowledge." *Educational Researcher* 41 (1): 26–28. doi:10.3102/0013189X11431010.

Van Keer, H., and J. P. Verhaeghe. 2005. "Comparing Two Teacher Development Programs for Innovating Reading Comprehension Instruction with Regard to Teachers' Experiences and Student Outcomes." *Teaching and Teacher Education* 21 (5): 543–562. doi:10.1016/j.tate.2005.03.002.

Van Veen, K., R. Zwart, and J. Meirink. 2012. "What Makes Teacher Professional Development Effective? A Literature Review." In *Teacher Learning That Matters: International Perspectives*, edited by M. Kooy and K. van Veen, 3–21. New York: Routledge.

Vanblaere, B., and G. Devos. 2016. "Exploring the Link between Experienced Teachers' Learning Outcomes and Individual and Professional Learning Community Characteristics." *School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice*. 27 (2): 205–227. doi:10.1080/09243453.2015.1064455.

Voerman, L., P. C. Meijer, F. Korthagen, and R. J. Simons. 2015. "Promoting Effective Teacher-feedback: From Theory to Practice through a Multicomponent Trajectory for Professional Development." *Teachers and Teaching* 21 (8): 990–1009. doi:10.1080/13540602.2015.1005868.

Vogt, F., and M. Rogalla. 2009. "Developing Adaptive Teaching Competency through Coaching." *Teaching and Teacher Education* 25 (8): 1051–1060. doi:10.1016/j.tate.2009.04.002.

Walker, A., M. Recker, L. Ye, M. B. Robertshaw, L. Sellers, and H. Leary. 2012. "Comparing Technology-related Teacher Professional Development Designs: A Multilevel Study of Teacher and Student Impacts." *Educational Technology Research and Development* 60 (3): 421–444. doi:10.1007/s11423-012-9243-8.

Wallace, M. 2009. "Making Sense of the Links: Professional Development, Teacher Practices, and Student Achievement." *The Teachers College Record* 111: 573–596. http://www.tcrecord.org/Content.asp?ContentID=15224.

Wasik, B. A., and A. H. Hindman. 2011. "Improving Vocabulary and Pre-literacy Skills of At-risk Preschoolers through Teacher Professional Development." *Journal of Educational Psychology* 103 (2): 455–469. doi:10.1037/a0023067.

Wayne, A. J., K. S. Yoon, P. Zhu, S. Cronen, and M. S. Garet. 2008. "Experimenting with Teacher Professional Development: Motives and Methods." *Educational Researcher* 37 (8): 469–479. doi:10.3102/0013189X08327154.

Willemse, T. M., G. ten Dam, F. Geijsel, L. van Wessum, and M. Volman. 2015. "Fostering Teachers' Professional Development for Citizenship Education." *Teaching and Teacher Education* 49: 118–127. doi:10.1016/j.tate.2015.03.008.

Zhang, M., M. Lundeberg, M. J. Koehler, and J. Eberhardt. 2011. "Understanding Affordances and Challenges of Three Types of Video for Teacher Professional Development." *Teaching and Teacher Education* 27 (2): 454–462. doi:10.1016/j.tate.2010.09.015.

Zwart, R. C., T. Wubbels, T. Bergen, and S. Bolhuis. 2009. "Which Characteristics of a Reciprocal Peer Coaching Context Affect Teacher Learning as Perceived by Teachers and Their Students?" *Journal of Teacher Education* 60 (3): 243–257. doi:10.1177/0022487109336968.

Appendix

	Authors	Publication type	PD outcomes situated in model Desimone	Measurement instruments (data collection)	Measurements included appendix?
1.	Abuhmaid (2011)	General impact study	Effective intervention features, Teacher quality, Teacher instruction	Interviews, questionnaires, direct classroom observations, field notes of classroom practices	Measurement instruments are not included in the appendix
2.	Antoniou and Kyriakides (2013)	Specific impact study	Teacher instruction, Student learning outcomes	Self-developed tests, student background characteristic questionnaire, time spent doing homework and time spent on private tuition, teacher characteristic questionnaire, classroom observations	Instruments are not included in the study. For the observation instrument, authors refer to previously published work
3.	Bakkenes, Vermunt, and Wubbels (2010)	Specific impact study	Teacher quality, Teacher instruction	Self-report through digital (structured) writing logs	Appendix A includes writing the digital learning log; tables include information on the categorisation of the writing logs
4.	Bierman et al. (2008)	Specific impact study	Student learning outcomes	Multi-method, multi-measure assessment battery including child assessments, teacher ratings, parent ratings and direct observations	The used tests are referred to but are not included in the study
5.	Boston (2013)	Specific impact study	Teacher quality	Assessment of teachers' knowledge of cognitive demands of mathematical tasks; videotaped discussions and written artefacts from the PD session	A sample task from the data sort task is provided in Figure 2
6.	Buczynski and Hansen (2010)	Specific impact study	Effective intervention features, Teacher quality, Teacher instruction, Student learning outcomes	Pre–post subject matter tests, teachers surveys, classroom observations, standardised and self-developed tests	Pre–post subject matter test for teachers are not included; Table 2 includes the items that were included in the end-of-year teacher survey; for the coding of the videos, the authors refer to other authors but no extended information on this is included; student achievement test is not included
7.	Butler et al. (2004)	General impact study	Teacher quality, Teacher instruction, Student learning outcomes	School visit notes, semi-structured observations of classroom instruction, teachers' reflection forms, summary notes from all-schools meetings, semi-structured teacher interviews, year-end interviews	Interviews are coded in relation with the research questions the authors proposed. No explicit coding protocol is included
8.	Chamberlin (2005)	Specific effect study	Teacher instruction	Individual students' Thinking Sheet (synthesis of the information in their notes on students' thinking and students' work), consensus students' thinking, teachers' collective interpretations while creating these consensus sheets	In the 'notes' section, website links are included in which the MEA's can be found. No examples are provided of students' thinking sheets. 'Interaction-patterns' are generated inductively and described by the authors
9.	Cheng and So (2012)	Specific effect study	Teacher quality, Teacher instruction	Semi-structured interviews, lesson observation reports, school visit reports, final documents reporting the support team's support and outcomes	Figure 2 represents the main coding categories by which the data are coded

10.	Cohen and Hill (2000)	General effect study	Teacher instruction, Student learning outcomes, School organisational conditions	Survey that probed teachers' familiarity with learning reform ideas, opportunities to learn about improved math instructions and reports of their mathematics teaching	In Tables 1, 2, 3 and 4, survey-items are provided. The note section provides some more (general) information on the used instruments (e.g. notes 12, 15, 21 and 26)
11.	Consuegra (2015)	Specific effect study	Teacher instruction, student learning outcomes, student characteristics	Questionnaire, video-recorded classroom observations	The instruments are not included in the appendix. For the questionnaire, they refer to a previously published questionnaire and provide a sample item in the text. The classroom observation schedule and coding procedure are included in Appendix 2 of the dissertation
12.	De Vries, van de Grift, and Jansen (2014)	General effect study	Teacher quality	Online survey	Questionnaire items are included in the table
13.	Dedousis-Wallace et al. (2014)	Specific effect study	Teacher quality, Effective intervention features	Vignettes to measure teacher self-predicted response to bullying situations; herein, teachers were provided with five scenarios, afterwards they had to answer different scales	Questionnaire items are included in the table
14.	Desimone et al. (2002)	General effect study	Effective intervention features, Teacher instruction	Surveys (self-reports)	Examples of specific teaching practices are provided in Table A1/A2. Descriptive information is provided on the elements included in the longitudinal teacher survey. The survey itself is not included in the appendix
15.	Domitrovich et al. (2009)	Specific effect study	Teacher quality, Teacher instruction, Student learning outcomes	Classroom observations, tests and questionnaires	The authors refer to previously published instruments that are not included in the appendix
16.	Doppelt et al. (2009)	Specific effect study	Student learning outcomes, Effective intervention features	Pre–post self-developed knowledge test, qualitative analyses of video from teacher workshops, informal observations of class activities	In the appendix, an example is provided of the knowledge test's items. No full instruments were included into this study
17.	Ekanayake and Wishart (2014)	Specific effect study	Teacher quality, Teacher instruction, Effective intervention features	Observation and field notes	The codebook is not included in the appendix
18.	Ermeling (2010)	Specific effect study	Effective intervention features, Teacher instruction	Extensive notes from face-to-face meetings, each classroom lesson was videotaped and coded, specific observations were recorded	No separate coding instrument is provided in the study (only textual description of the different categories)
19.	Fishman et al. (2003)	Specific effect study	Teacher quality, Teacher instruction, Student learning outcomes	Opinionnaires (i.e. traditional survey-based instruments) tape-recordings of focus groups, videotapes and notes of all PD activities, notes from classroom observations, pre-and post-tests (written assessment instruments with multiple-choice and free response items)	instruments are not included in the study
20.	Franke et al. (2001)	Specific effect study	Teacher instruction	Classroom observations, interviews	Textual description on how the interview was conducted; Table 2 documents the levels of engagement with children's mathematical thinking (used for coding)

(*Continued*)

Appendix. (*Continued*)

	Authors	Publication type	PD outcomes situated in model Desimone	Measurement instruments (data collection)	Measurements included appendix?
21.	Garet et al. (2008)	Specific effect study	Teacher quality, Teacher instruction, Student learning outcomes	Survey, multiple-choice and short constructed response assessment created classroom observations, student achievement tests	More information on the used instruments used for student and teacher learning are provided from Appendix D on (D: includes not the full instrument, only information on statistical analyses concerning the instrument).>E: items used to create different scales (e.g. explicit instruction, independent study activity and/or differentiated instruction) are explicitly described
22.	Garet et al. (2001)	General effect study	Teacher quality, Teacher instruction	Survey	The instrument (items) is described in the text; the full instrument is not included in the appendix
23.	Goldschmidt and Phelps (2010)	Specific effect study	Teacher quality	Survey	The authors refer to previously published work
24.	Hofman and Dijkstra (2010)	Specific effect study	Effective intervention features, Teacher quality, Teacher instruction, Student learning outcomes, Personal characteristics	Internet questionnaire	The full instrument (internet questionnaire) is not included in the study, questionnaire items can be found in the Tables (2–5)
25.	Ingvarson, Meiers, and Beavis (2005)	General effect study	Effective intervention features, Teacher quality, Teacher instruction, Student learning outcomes, Personal characteristics	Survey	The full instrument is not included in appendix but all items are described in the study itself
26.	James and McCormick (2009)	General effect study	Teacher quality, Teacher instruction, School organisational conditions, Student learning outcomes	Video-recordings of lessons, teacher interviews, survey data, performance data of pupils from national databases, self-developed questionnaire	Video-coding protocols are not provided. Survey is not included. The 84-item self-developed questionnaire is textually described (the questionnaire itself is not fully provided in the article)
27.	Kazemi and Franke (2004)	Specific effect study	Teacher instruction	Transcripts from audio-recordings during work-group meetings, written teacher reflections, copies of student work shared by teachers, end-of- year teacher interviews	The coding categories for the analysis of the audio-transcripts are included in Table 4
28.	Kiemer et al. (2015)	Specific effect study	Teacher instruction, Student learning outcomes, Effective intervention features	Questionnaires: scales adapted from PISA, self-developed scales	For the vignettes, they refer to a previously published work; however, descriptive information in the text is provided also for the scale, previous work is cites, some textual examples are given
29.	Lee et al. (2004)	Specific effect study	Teacher quality, Teacher instruction	Focus group interviews, self-developed questionnaire, classroom observations	Protocol for focus group interview codes and the questionnaire (items) are not separately included. The descriptive codes used for the focus group data can be found in the text. The self-develop observation scale is textually described (1029)

30.	Levenson and Gal (2013)	Specific effect study	Teacher quality, Teacher instruction	Questionnaires, interviews	Some example items of the questionnaire are textually described, the instrument itself is not included in the appendix; regarding the semi-structured interviews, the questions are provided textually
31.	Levine and Marcus (2010)	Specific effect study	Effective intervention features, Teacher quality, Teacher instruction	Field notes of collaborative meetings document collection, transcriptions taped meeting, interviews, classroom observation	Appendix A represents an excerpt of codes; Appendix B represents the critical friend protocol
32.	Liu (2013)	Specific effect study	School organisational conditions, Teacher instruction, Teacher quality	Instructional observations and focus group interviews	Measurement instruments are not included in the appendix
33.	McCutchen et al. (2002)	Specific effect study	Teacher quality, Teacher instruction, Student learning outcomes	Survey, tests, notes from classroom observations	For teacher knowledge and student learning, already existing tests are used and referred to. The self-developed coding scheme for teaching practice is textually described
34.	Mikami et al. (2011)	Specific effect study	Teacher instruction	Classroom videotapes, self-report measures	None of the instruments used are included in Appendix. The authors refer to previously published measurement instruments (e.g. CLASS observation system)
35.	Morge, Toczek, and Chakroun (2010)	Specific effect study	Teacher instruction, Student learning outcomes	Recordings of teachers' sessions, an open question and a test questionnaire for the pupils	The pupils' questionnaire is presented in Appendices A and B
36.	Morrison (2014)	Specific effect study	Teacher quality, Teacher instruction	Survey, reflection papers, questionnaire, classroom observation	In the study, they refer to other studies concerning the coding of the videos and written self-reports. No detailed information is provided
37.	Norton and McCloskey (2008)	Specific effect study	Teacher quality	Audio-recordings, classroom observations, follow-up interview	Table 1 represents an example of a classroom observation matrix. The interview coding scheme is not included in the manuscript. Themes emerging from the analyses are textually described
38.	Opfer and Pedder (2011)	General effect study	Effective intervention features, School organisational conditions	Survey	The instrument is not included in the appendix. The different sections are textually described and items are provided in Tables 1 and 2
39.	Pehmer, Gröschner, and Seidel (2015)	Specific effect study	Student learning outcomes, Personal (student) characteristics	Questionnaires	The scales are not included in the appendix. They refer to previously published work
40.	Ponte et al. (2004)	Specific effect study	Teacher quality	Structured logbooks, recording of the meetings, supplementary interviews and documents	Questions integrated into the logbooks; standardised questions for fragment analysis are textually described under 'data-collection'; Table 2 documents on the categories as reported by the teachers themselves at the end of the project
41.	Saka (2013)	Specific effect study	Effective intervention features, Teacher quality, Student characteristics	Questionnaire, interviews	The questionnaire scales are referred to in the text. Subscales of these instruments are textually described. The interview protocol is not included in the study. For the coding procedures, the authors refer to the Loucks–Horsley model. No further, more detailed, information on this model is provided

(*Continued*)

Appendix. (*Continued*)

	Authors	Publication type	PD outcomes situated in model Desimone	Measurement instruments (data collection)	Measurements included appendix?
42.	Saxe, Gearhart, and Nasir (2001)	Specific effect study	Student learning outcomes	Test	The items of the self-developed student test are listed in Appendix A
43.	Supovitz and Turner (2000)	General effect study	Teacher quality, Teacher instruction, School organisational conditions	Survey	In the appendix, survey questions and scale used to create science composites are presented
44.	Teemant, Wink, and Tyra (2011)	Specific effect study	Effective intervention features, teacher instruction	Observations	The authors refer to previously published measurement instruments
45.	Tienken and Achilles (2003)	Specific effect study	Teacher instruction, Student learning outcomes	Narrative writing picture-prompts, interviews, classroom observations	Coding scheme for classroom observations (~instructional mode of teachers) is not included into the study. The scoring rubric for the writing tasks is textually described (broad categories); no appendix or examples are provided
46.	Van den Bergh, Ros, and Beijaard (2015)	Specific effect study	Teacher instruction	Videotaped observations of teachers' feedback behaviour and participation in the video interaction training, written self-reports	Measurement instruments are referred to but not included in the article
47.	Voerman et al. (2015)	Specific effect study	Teacher instruction	Videotaped lessons	The authors refer to an already developed observation scheme in a previous study
48.	Vogt and Rogalla (2009)	Specific effect study	Teacher quality, Teacher instruction, Student learning outcomes	Vignettes, video tests, scientific literacy tests	Table 2 represents the criteria and indicators for the scoring of the vignettes. Table 3 represents an example from the rating manual for the video test coding
49.	Walker et al. (2012)	Specific effect study	Effective intervention features, Teacher quality, Teacher instruction, Student learning outcomes	Teacher survey, web usage data, questionnaires	Instruments are not included in the appendix. For the teacher survey, the authors refer to previously published instruments from others and themselves. The rubric to see whether teachers' activities are PBL aligned is included in appendix. Some items of the student questionnaire are textually described, though the survey is not included in appendix
50.	Wallace (2009)	General effect study	Teacher instruction, Student learning outcomes	Scales constructed from self-report data	No instruments are documented in the study
51.	Wasik and Hindman (2011)	Specific effect study	Effective intervention features, Teacher quality, Teacher instruction, Student learning outcomes	Observations, checklists, tests	Measurement instruments are not included in the appendix, though the authors refer to previously published measurement instruments
52.	Willemse et al. (2015)	Specific effect study	Teacher quality	Interviews	Categories are represented in Table 2
53.	Zhang et al. (2011)	Specific effect study	Effective intervention features	Survey, reports, discussions, focus group interviews, phone interviews, video-recordings, chart and note collections	No instruments are included in Appendix. A coding rubric is developed and textually described
54.	Zwart et al. (2009)	Specific effect study	Teacher instruction	Questionnaires, self-reports, digital diaries	The questionnaire for peer coaching context characteristics is not included in the study. See Figure 2: overview of the coding procedure of the digital diaries